164149C - Wijethunge O.U.M

**Individual contribution**

I was assigned to detect spam comments from YouTube comments section and make a legitimate comments details list. I have studied number of research papers and watched some online tutorials for find an appropriate approaches and methods for implement my assigned task and I have tested few approaches and chose a better approach for my task. I downloaded a YouTube comments details file and a YouTube content details file as my initial input data sets. Mainly I used YouTube data API, YouTube dll and google speech recognition as supporting libraries for make the datasets. I used cosine similarity approach for find the related comments from whole comments and also I used tf-idf values for create my comments vectors for further calculations. I used number of python libraries for preprocessing and measure extracted features data. I have extracted number of features such as Similarity between content and comment, Repeating duplicates comments which are posted by same user in a same video, Similarity between comment - comment (surrounding comments), Word count of a comment, Words duplication ratio of a comment, Number of sentences of a comment, Number of punctuation marks in a comment , Black words ration words in a comment, Length of a comment, Availability of url, Availability of YouTube url , Availability of phone numbers in a comment , availability of period marks sequence in a comments, Time gap of Video posted date time and comment posted date time from seconds.by using comment details and commenter details and made a csv file for input to classification algorithm. Overall I used sklearn, pandas, numpy, scipy, matplotlib, seaborn, tensor flow as main libraries for build my sub module.

I have trained and tested nine machine learning classification algorithm such as logistic regression, artificial neural network, random forest, Naive Bayes, k-nearest neighbor, decision tree, support vector machine, gradient boost and extremely gradient boost for find the well suited model for predicts comments spam from whole comments section in YouTube. after evaluating models. I chose gradient boosting model as my final model. gradient boosting model use ensemble techniques for predicts and it has number of decision trees sequentially and it use a voting classifier for give a better output. I did hyper parameter tuning to avoid model getting overfitting issue and made a well generalized model. after setting up my sub module architecture I made a legitimate comments detail data frame (no spam comments) as my final output to send the next sub module.

**Further work**

Spammers’ behaviors are change day by day therefor spam filtering techniques also should update according to the new spammers’ behaviors such as spammers use new words and edit their words. Hence spam filtering sub module should be able to catch new black words for maintain the productivity of it for a long time. Therefor I suppose to implement an interactive training model for catch those new features to enhance the quality of product.